2. The Perceptron Model

2.1 Introduction

The name of this chapter comes from the theory of neural networks. An accessible introduction to neural networks is provided in [83], but what these are is not relevant to our purpose, which is to study the underlying mathematics. Roughly speaking, the basic problem is as follows. What “proportion” of \( \Sigma_N = \{-1,1\}^N \) is left when one intersects this set with many random half-spaces? A natural definition for a random half-space is a set \( \{ x \in \mathbb{R}^N ; x \cdot v \geq 0 \} \) where the random vector \( v \) is uniform over the unit sphere of \( \mathbb{R}^N \). More conveniently one can consider the set \( \{ x \in \mathbb{R}^N ; x \cdot g \geq 0 \} \), where \( g \) is a standard Gaussian vector, i.e. \( g = (g_i)_{i \leq N} \), where \( g_i \) are independent standard Gaussian r.v.s. This is equivalent because the vector \( g/\|g\| \) is uniformly distributed on the unit sphere of \( \mathbb{R}^N \). Consider now \( M \) such Gaussian vectors \( g_k = (g_{i,k})_{i \leq N}, k \leq M \), all independent, the half-spaces

\[ U_k = \{ x ; x \cdot g_k \geq 0 \} = \left\{ x, \sum_{i \leq N} g_{i,k} x_i \geq 0 \right\}, \]

and the set

\[ \Sigma_N \cap \bigcap_{k \leq M} U_k. \tag{2.1} \]

A given point of \( \Sigma_N \) has exactly a 50% chance to belong to \( U_k \), so that

\[ \mathbb{E} \operatorname{card} \left( \Sigma_N \cap \bigcap_{k \leq M} U_k \right) = 2^{N-M}. \tag{2.2} \]

The case of interest is when \( N \) becomes large and \( M \) is proportional to \( N \), \( M/N \rightarrow \alpha > 0 \). A consequence of (2.2) is that if \( \alpha > 1 \) the set (2.1) is typically empty when \( N \) is large, because the expected value of its cardinality is \( \ll 1 \). When \( \alpha < 1 \), what is interesting is not however the expected value (2.2) of the cardinality of the set (2.1), but rather the typical value of this cardinality, which is likely to be smaller. Our ultimate goal is the computation of this typical value, which we will achieve only for \( \alpha \) small enough.

A similar problem was considered in (0.2) where \( \Sigma_N \) is replaced by the sphere \( S_N \) of center 0 and radius \( \sqrt{N} \). The situation with \( \Sigma_N \) is usually

called the binary perceptron, while the situation with $S_N$ is usually called the spherical perceptron. The spherical perceptron will motivate the next chapter. We will return to both the binary and the spherical perceptron in Volume II, in Chapter 8 and Chapter 9 respectively. Both the spherical and the binary perceptron admit another popular version, where the Gaussian r.v.s $g_{i,j}$ are replaced by independent Bernoulli r.v.s (i.e. independent random signs), and we will also study these. Thus we will eventually investigate a total of four related but different models. It is not very difficult to replace the Gaussian r.v.s by random signs; but it is very much harder to study the case of $\Sigma_N$ than the case of the sphere.

Research Problem 2.1.1. (Level 3!) Prove that there exists a number $\alpha^*$ and a function $\varphi : [0, \alpha^*) \to \mathbb{R}$ with the following properties:

1-If $\alpha > \alpha^*$, then as $N \to \infty$ and $M/N \to \alpha$ the probability that the set (2.1) is not empty is at most $\exp(-N/K(\alpha))$.

2-If $\alpha < \alpha^*$, $N \to \infty$ and $M/N \to \alpha$, then

$$\frac{1}{N} \log \text{card} \left( \Sigma_N \cap \bigcap_{k \leq M} U_k \right) \to \varphi(\alpha)$$

(2.3)
in probability. Compute $\alpha^*$ and $\varphi$.

This problem is a typical example of a situation where one expects “regularity” as $N \to \infty$, but where it is unclear how to even start doing anything relevant. In Volume II, we will prove (2.3) when $\alpha$ is small enough, and we will compute $\varphi(\alpha)$ in that case. (We expect that the case of larger $\alpha$ is much more difficult.) As a corollary, we will prove that there exists a number $\alpha_0 < 1$ such that if $M = \lfloor \alpha N \rfloor$, $\alpha > \alpha_0$, then the set (2.1) is typically empty for $N$ large, despite the fact that the expected value of its cardinality is $2^{N-M} \gg 1$.

One way to approach the (very difficult) problem mentioned above is to introduce a version “with a temperature”. We observe that if $x \geq 0$ we have $\lim_{\beta \to \infty} \exp(-\beta x) = 0$ if $x > 0$ and $= 1$ if $x = 0$. Using this for $x = \sum_{k \leq M} 1_{\{\sigma \notin U_k\}}$ where $\sigma \in \Sigma_N$ implies

$$\text{card} \left( \Sigma_N \cap \bigcap_{k \leq M} U_k \right) = \lim_{\beta \to \infty} \sum_{\sigma \in \Sigma_N} \exp\left( -\beta \sum_{k \leq M} 1_{\{\sigma \notin U_k\}} \right),$$

(2.4)
so that to study (2.3) it should be relevant to use the Hamiltonian

$$-H_{N,M}(\sigma) = -\beta \sum_{k \leq M} 1_{\{\sigma \notin U_k\}}.$$  

(2.5)

If one can compute the corresponding partition function (and succeed in exchanging the limits $N \to \infty$ and $\beta \to \infty$), one will then prove (2.3).

More generally, we will consider Hamiltonians of the type